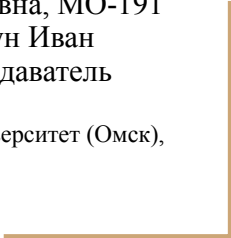


АЛГОРИТМ
ФОРМИРОВАНИЯ
КОМПЕТЕНТНОСТНОГО
ПРОФИЛЯ СТУДЕНТА
С ПОМОЩЬЮ МЕТОДОВ
МАШИННОГО ОБУЧЕНИЯ

Автор: Катюшкина Алина Валерьевна, МО-191
Научный руководитель: Шарун Иван
Владимирович, старший преподаватель

Омский государственный технический университет (Омск),
Россия



Актуальность

Задачами *NER (Named Entity Recognition)* являются автоматическое обнаружение и классификация именованных сущности в тексте, такие как имена людей, организации, места и другие сущности. Это позволяет упростить процесс анализа больших объемов текстовой информации и повысить точность извлечения нужных данных.

В контексте данного исследования автоматически были извлечены дескрипторные характеристики студента из текстовых источников, с помощью методов *NLP (Natural Language Processing)* позволяет, что упрощает процесс оценки его знаний и умений. Это актуально для образовательных учреждений и компаний, занимающихся подбором персонала.

Исследования

Примеры аналогичных исследований:

- АЛГОРИТМ ФОРМИРОВАНИЯ ДЕСКРИПТОРНОЙ СУЩНОСТИ ВАКАНСИИ ПРОФЕССИОНАЛЬНОЙ ОБЛАСТИ ДЕЯТЕЛЬНОСТИ
(А.В. Теремшенко, О.Н. Канева)
- РАЗРАБОТКА АЛГОРИТМА НАХОЖДЕНИЯ СЕМАНТИЧЕСКОЙ БЛИЗОСТИ МЕЖДУ КОМПЕТЕНЦИЯМИ
(О. Н. Канева, К. А. Акимова, И. В. Шарун)
- РАЗРАБОТКА АЛГОРИТМА ФОРМИРОВАНИЯ ДЕСКРИПТОРНОЙ МОДЕЛИ ДЛЯ УЧЕБНОЙ ДИСЦИПЛИНЫ
(О. Н. Канева, М. Н. Вагипова, И. В. Шарун)
- РАЗРАБОТКА ПОДХОДА ДЛЯ ПОЛУЧЕНИЯ АВТОМАТИЗИРОВАННОГО РЕШЕНИЯ ФОРМИРОВАНИЯ УЧЕБНОГО ПЛАНА
(А.В. Зыкина, О.Н. Канева, В.В. Мунько)

Понятие дескрипторной сущности

Дескрипторная сущность – это набор характеристик, описывающих конкретный объект или явление.

$$K = \langle S^K, D^K, T^K \rangle$$

S - формулировка компетенции

Множество дескрипторов

$$D = \{D_1, D_2, \dots, D_m\}$$

$$D_j = \langle S_{D_j}, V^j \rangle, j = \overline{1, m}$$

$$V^j = \{v_1^j, v_2^j, \dots, v_{p_j}^j\}$$

Множество терминов

$$T = \{T_1, T_2, \dots, T_n\}$$

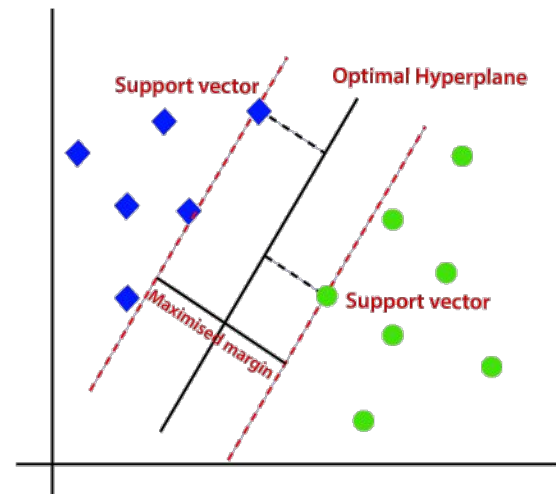
$$T_i = \langle S_{D_i}, W^i \rangle, i = \overline{1, n}$$

$$W^i = \{W_1^i, W_2^i, \dots, W_m^i\}$$

$$W_j^i = \{w_{j1}^i, w_{j2}^i, \dots, w_{jp_j}^i\}$$

Алгоритм *SVM*

Метод опорных векторов (*SVM*) – это алгоритм машинного обучения, который используется для классификации и регрессии. Он основан на поиске гиперплоскости, которая разделяет данные на два класса с максимальным зазором между ними.



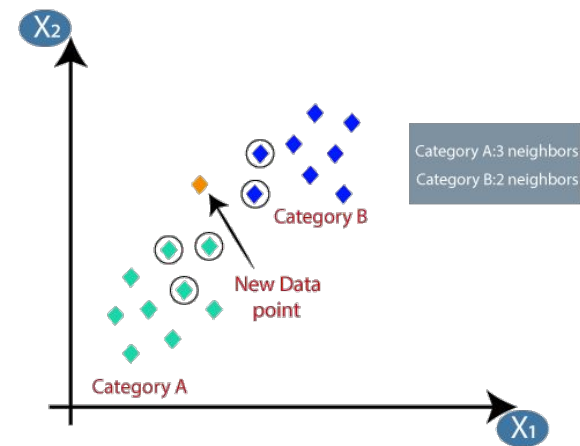
Деревья принятия решений

Деревья принятия решений – это алгоритм машинного обучения, который используется для классификации и регрессии. Он представляет собой древовидную структуру, где каждый узел представляет собой тест на значение одного из признаков, а каждая ветвь представляет собой возможный результат этого теста.



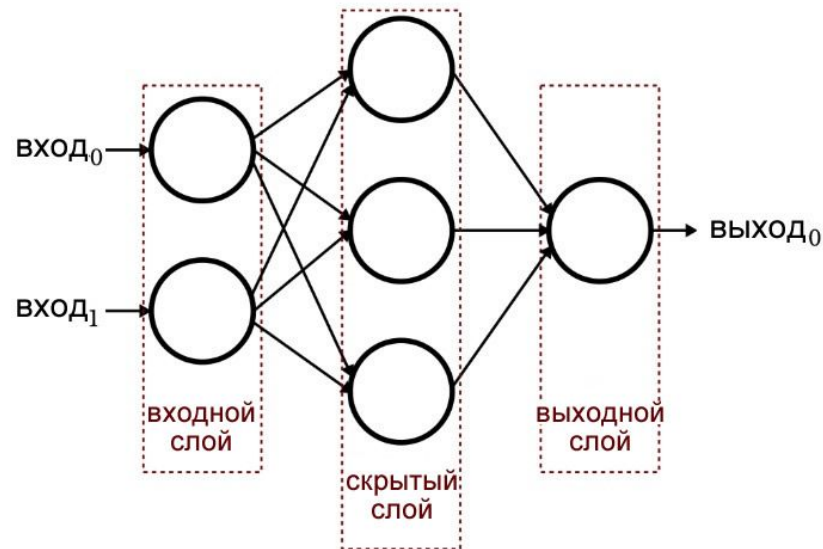
Метод k-ближайших соседей

Метод k-ближайших соседей (k -NN) – это алгоритм машинного обучения, который используется для классификации и регрессии. Он основан на принципе близости объектов и предполагает, что объекты, находящиеся близко друг к другу в пространстве признаков, имеют похожие метки классов.



Многослойный перцептрон

Многослойный перцептрон (*MLP*) – это алгоритм машинного обучения, который используется для классификации и регрессии. Он состоит из нескольких слоев нейронов, которые обрабатывают входные данные и создают выходные данные.



Исследованные алгоритмы и их результаты

Сводная таблица и список алгоритмов

Метрика\алгоритм	SVM	Деревья принятия решений	<i>k</i> -NN	MLP
recall	0.82	0.77	0.77	0.81
precision	0.95	0.79	0.87	0.94
AUC	0.91	0.88	0.88	0.91
F1	0.88	0.78	0.81	0.87

Распределение по дескрипторам

Посчитаем расстояние до гиперплоскости для всех данных, у которых это расстояние положительное, а значит относится к описанию студента, и нормируем его от 0 до 3. Если нормированное расстояние находится в промежутке от 0 до 1, то будем его относить его к дескриптору «знать», если от 1 до 2, то к «уметь», и в качестве веса возьмем это нормированное расстояние и отнимем 1, а если от 2 до 3, то к «владеть», и в качестве веса возьмем это нормированное расстояние и отнимем 2.

```
"descriptor_model": {  
  "владеть": [],  
  "знать": [  
    {  
      "text": "grain",  
      "value": 0.26757460006916184  
    },  
    {  
      "text": "trade",  
      "value": 0.6266948218751527  
    }  
  ],  
  "уметь": []  
},
```

Результаты

Разработанный алгоритм для формирования компетентностного профиля студента позволяет оценить уровень студента. Однако, не все дескрипторы были заполнены, что указывает на необходимость использования более объемного корпуса данных для более полного представления каждого значения дескриптора.

Спасибо за внимание